On the Berry–Esseen bound for the Student statistic

Iosif Pinelis[*]

Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
E-mail: ipinelis@mtu.edu

Abstract: New Berry–Esseen-type bounds, with explicit constant factors, for the distribution of the Student statistic and, equivalently, for that of the self-normalized sum of independent zero-mean random variables are obtained. These bounds are compared with the corresponding existing results.

Contents

1. Summary and discussion

Consider the self-normalized sum

$$T := \frac{S}{V},$$

where

$$S := \sum_1^n X_i, \quad V := \sqrt{\sum_1^n X_i^2},$$

and $X_1, \ldots, X_n$ are independent zero-mean random variables (r.v.'s). It is assumed that $T = 0$ on the event $\{V = 0\}$. For any $p \in (0, \infty)$, introduce also

$$\beta_p := \sum_1^n \mathsf{E}\,|X_i|^p \quad \text{and} \quad \tilde{\beta}_p := \sum_1^n \mathsf{E}\,|X_i^2 - \mathsf{E}\,X_i^2|^{p/2},$$

assuming that $0 < \beta_3 < \infty$ (and hence $0 < \beta_2 < \infty$).

Let Φ be the standard normal distribution function.

Theorem 1.1. *One has*

$$|\,\mathsf{P}(T \leqslant z) - \Phi(z)| \leqslant A_3 \frac{\beta_3}{\beta_2^{3/2}} + A_4 \frac{\tilde{\beta}_4^{1/2}}{\beta_2} + A_6 \frac{\tilde{\beta}_6}{\beta_3^3 \beta_2^{3/2}} \tag{1.1}$$

for all $z \in \mathbb{R}$ and for all triples $\tau := (A_3, A_4, A_6)$ of absolute constants belonging to the set $\mathcal{T} := \{\tau_1, \ldots, \tau_4\}$ of triples, where

$$\begin{aligned}
\tau_1 &:= (1.61, 1.60, 1.20), \\
\tau_2 &:= (2.01, 1.02, 0.61), \\
\tau_3 &:= (11.38, 11.02, 11.78 \times 10^{-6}), \\
\tau_4 &:= (1.34, 125377, 1.049 \times 10^6).
\end{aligned}$$

The triple $\tau_1 = (1.61, 1.60, 1.20)$ of the constant factors A_3, A_4, A_6 was obtained trying to minimize the maximum $A_3 \vee A_4 \vee A_6$ of the constants; for details, see the proof (in Section 2) of Theorem 1.1 and especially the table at the end of that proof. The triple τ_3 was obtained trying to minimize the effect of the 6th-order moments of the X_i's. The triple τ_4 was designed to work best when $\tilde{\beta}_4$ and $\tilde{\beta}_6$ are very small, that is, when the distribution of each X_i is close to the symmetric distribution on a symmetric two-point set. The triples τ_2, τ_3, τ_4 will be used in this paper to compare the upper bound in (1.1) with one due to Shao [16].

In the i.i.d. case, that is, when the r.v.'s $X_1, \ldots, X_n$ are independent copies of a r.v. X, one can improve the values A_3, A_4, A_6 of the absolute constants in (1.1); at that, let us assume without loss of generality that

$$\mathsf{E}\,X^2 = 1.$$

Introduce

$$\rho_3 := \mathsf{E}\,|X|^3, \quad \rho_4 := \sqrt{\mathsf{E}(X^2-1)^2}, \quad \rho_6 := \frac{\mathsf{E}\,|X^2-1|^3}{\mathsf{E}\,|X|^3}. \tag{1.2}$$

Theorem 1.2. *If $X, X_1, \ldots, X_n$ are i.i.d. r.v.'s with $\mathsf{E}\,X = 0$, $\mathsf{E}\,X^2 = 1$, and $\mathsf{E}\,|X|^3 < \infty$, then*

$$|\,\mathsf{P}(T \leqslant z) - \Phi(z)| \leqslant \frac{A_3\rho_3 + A_4\rho_4 + A_6\rho_6}{\sqrt{n}} \tag{1.3}$$

for all $z \in \mathbb{R}$ and for all triples $\tau := (A_3, A_4, A_6)$ of absolute constants belonging to the set $\tilde{\mathcal{T}} := \{\tilde{\tau}_{1,1}, \ldots, \tilde{\tau}_{4,1}\}$ of triples, where

$$\begin{aligned}
\tilde{\tau}_{1,1} &:= (1.53, 1.52, 1.34), \\
\tilde{\tau}_{1,2} &:= (1.61, 1.60, 1.02), \\
\tilde{\tau}_{2,2} &:= (1.96, 1.02, 0.52), \\
\tilde{\tau}_{2.1,1} &:= (1.96, 0.99, 0.63), \\
\tilde{\tau}_{3,1} &:= (10.94, 9.40, 11.06 \times 10^{-6}), \\
\tilde{\tau}_{4,1} &:= (1.25, 8140, 92437);
\end{aligned}$$

here, for each $i = 1, 2, 3, 4$, the triples $\tilde{\tau}_{i,j}$ are to be compared with the triple τ_i in Theorem 1.1, with the same i.

For $n \geqslant 2$, the Student statistic

$$t := \frac{\overline{X}\sqrt{n}}{\sqrt{\frac{1}{n-1}\sum_1^n (X_i - \overline{X})^2}},$$

where $\overline{X} := \frac{1}{n}\sum_1^n X_i$, can be expressed as a monotonic transformation of the self-normalized sum T:

$$t = \sqrt{\frac{n-1}{n}}\,\frac{T}{\sqrt{1 - T^2/n}}. \tag{1.4}$$

Therefore, one immediately has

Corollary 1.3. *Theorems 1.1 and 1.2 hold if $\mathsf{P}(T \leqslant z) - \Phi(z)$ is replaced there by $\mathsf{P}(t \leqslant z) - \Phi_n(z)$, where*

$$\Phi_n(z) := \Phi\Big(\frac{z}{\sqrt{1 + (z^2-1)/n}}\Big). \tag{1.5}$$

A Berry-Esseen type of bound of the optimal order for the Student statistic of i.i.d. X_i's was obtained in 1996 by Bentkus and Götze [2], using a Fourier transformation method. This was extended to the non-i.i.d. case by Bentkus, Bloznelis, and Götze [1], whose result can be rewritten as follows:

$$|\,\mathsf{P}(t \leqslant z) - \Phi\big(z\sqrt{\tfrac{n}{n-1}}\big)| \leqslant C_2\gamma_2 + C_3\gamma_3, \tag{1.6}$$

where C_2 and C_3 are absolute constants,

$$\gamma_2 := \frac{1}{\beta_2} \sum_1^n \mathsf{E}\, X_i^2\, \mathbf{I}\Big\{|X_i| > \frac{\sqrt{\beta_2}}{2}\Big\}, \quad \gamma_3 := \frac{1}{\beta_2^{3/2}} \sum_1^n \mathsf{E}\, |X_i|^3\, \mathbf{I}\Big\{|X_i| \leqslant \frac{\sqrt{\beta_2}}{2}\Big\}. \tag{1.7}$$

Note that $t \sim T$ as $n \to \infty$. The function Φ_n, defined by (1.5), may be considered as an improper distribution function, with the "impropriety" $1 - \big(\Phi_n(\infty) - \Phi_n(-\infty)\big) = 2\big(1 - \Phi(\sqrt{n})\big) \sim \sqrt{\frac{2}{\pi n}}\, e^{-n/2}$ for large n, which is much less than $\frac{1}{\sqrt{n}}$. If n is not very large, the tail probability $1 - \Phi_n(z)$ may be much greater than $1 - \Phi(z)$, which appears to correspond qualitatively to the fact that the tail of the Student distribution is significantly heavier than the standard normal tail when the number of degrees of freedom (d.f.) is not large. This heuristics appears to be confirmed by Figure 1, for $n = 10$; the pictures for $n = 5$ and $n = 20$ look quite similarly.
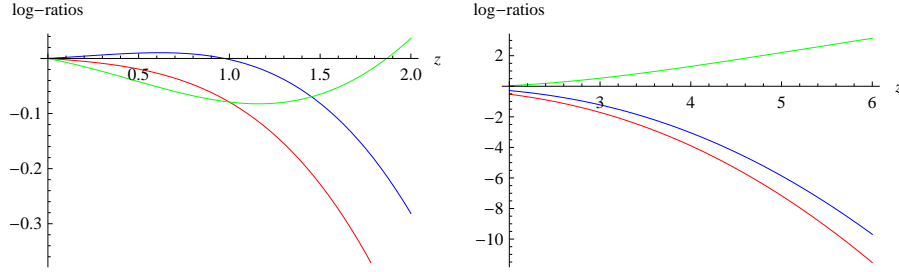

FIG 1. *Logarithms of the ratios of the tail functions $1 - \Phi(\cdot)$ (red), $1 - \Phi\big(\cdot\sqrt{\frac{n}{n-1}}\big)$ (blue), and $1 - \Phi_n(\cdot)$ (green) to the tail function of the Student distribution with $n - 1$ d.f.*

It appears that on the interval $[1.5, \infty)$ the tail function $1 - \Phi_n(\cdot)$ is closer to that of the Student distribution than the tail functions $1 - \Phi(\cdot)$ and $1 - \Phi\big(\cdot\sqrt{\frac{n}{n-1}}\big)$ are. So, while the method of the proof (given in Section 2) appears to allow one to obtain analogs of Theorems 1.1 and 1.2 for $\mathsf{P}(t \leqslant z) - \Phi(z)$ in place of $\mathsf{P}(T \leqslant z) - \Phi(z)$ or $\mathsf{P}(t \leqslant z) - \Phi_n(z)$, such analogs will not be pursued here.

Anyway, the following proposition shows that $\Phi_n(z)$ differs from $\Phi(z)$ by much less than $1/\sqrt{n}$, uniformly in $z \in \mathbb{R}$.

Proposition 1.4. *For all $n > 1$ and $z \in \mathbb{R}$*

$$|\Phi(z) - \Phi_n(z)| < \frac{C}{n-1}, \quad where \tag{1.8}$$

$$C := \Big(k - \frac{1}{2}\Big)e^{-k}\sqrt{\frac{k}{\pi}} = 0.162\dots \quad and \quad k := 1 + \frac{\sqrt{3}}{2};$$

this constant factor, C, is the best possible in (1.8).

One may be concerned that it is more natural to compare the distribution function of the statistic t $\big($as in (1.4), for general zero-mean X_i's$\big)$, not with Φ or Φ_n, but with the distribution function (say F_{n-1}) of Student's distribution with $n-1$ d.f. — that is, with the distribution function of the statistic t for i.i.d. standard normal X_i's. However, as shown in [10],

$$|F_{n-1}(z) - \Phi(z)| < \frac{\tilde{C}}{n-1} \quad \text{with} \quad \tilde{C} = 0.158\dots$$

for all $n \geqslant 5$ and $z \in \mathbb{R}$. Therefore and in view of Proposition 1.4, $F_{n-1}(z)$ differs from $\Phi_n(z)$ by much less than $1/\sqrt{n}$, uniformly in $z \in \mathbb{R}$. Thus, Corollary 1.3 is quite relevant, notwithstanding the mentioned concern.

In the i.i.d. case, Nagaev [6, (1.18)] stated an inequality, which reads as follows (in the conditions of Theorem 1.2): for all $z \in \mathbb{R}$

$$|\, \mathsf{P}(T \leqslant z) - \Phi(z)| < \Big(4.4\, \mathsf{E}\, |X|^3 + \frac{\mathsf{E}\, X^4}{\mathsf{E}\, |X|^3} + \mathsf{E}\, |X^2 - 1|^3\Big) \frac{1}{\sqrt{n}}. \qquad (1.9)$$

However, there are a number of mistakes in the proof of (1.9) in [6]. It is also stated in [6], again in the i.i.d. case, that

$$|\, \mathsf{P}(T \leqslant z) - \Phi(z)| < \frac{36\, \mathsf{E}\, |X|^3 + 9}{\sqrt{n}}.$$

Using Stein's method, Shao [16] obtained a tighter and more general bound, also with explicit constants but without the i.i.d. assumption:

$$|\, \mathsf{P}(T \leqslant z) - \Phi(z)| \leqslant 10.2\gamma_2 + 25\gamma_3 \qquad (1.10)$$
$$\leqslant 25\beta_p/\beta_2^{p/2}$$

for all $p \in [2, 3]$, with the same γ_2 and γ_3 as in (1.7). More recently, a Berry–Esseen bound for T was obtained in [3] for i.i.d. standard normal X_i's by means of Malliavin calculus.

Let us compare the bounds in (1.1) and (1.10). At that, let us restrict the attention to i.i.d. r.v.'s $X, X_1, \dots, X_n$.

Consider first the case when X has a two-point zero-mean distribution, so that

$\mathsf{P}(X \in \{-a, b\}) = 1$ for some positive real numbers a and b; that is,

$$\mathsf{P}(X = b) = \frac{a}{a+b} = 1 - \mathsf{P}(X = -a).$$

This case appears especially interesting, as any zero-mean distribution can be represented as a mixture of two-point zero-mean distributions — see e.g. [15]. Without loss of generality, assume that $b \geqslant a$ and $ab = 1$. Then $b \geqslant 1$ and $\mathsf{E}\, X^2 = 1$, and hence the bound in (1.1) (with the triple $\tau = \tau_3$ of constants A_3, A_4, A_6) is no greater than $(11.38\rho_3 + 11.02\rho_4 + 11.78 \times 10^{-6}\, \rho_6)/\sqrt{n}$, where again the ρ_j's are as in (1.2), so that $\rho_3 = \frac{b^4+1}{b(b^2+1)}$, $\rho_4 = b - 1/b$, and $\rho_6 =$

$(b - 1/b)^3$. On the other hand, if $b > \sqrt{n}/2$, then the bound in (1.10) is no less than $10.2\frac{b}{b+1/b} \geqslant 5.1 > 1$. So, without loss of generality $b \leqslant \sqrt{n}/2$ and hence the bound in (1.10) equals $25\rho_3/\sqrt{n}$. Thus (preferably with the help of the Mathematica command Reduce or similar tools), one finds that the bound in (1.10) will be less than the bound in (1.1) only if $b > 469$, that is, only if the "asymmetry index" b/a is greater than $469^2 = 219961$; at that, the inequality $b \leqslant \sqrt{n}/2$ implies that n must be no less that $(2b)^2 > (2 \times 469)^2 = 879844$. One concludes that, for i.i.d. X_i's with a common two-point distribution, (1.1) is better than (1.10) unless both the sample size n and the asymmetry index are very large. Also, in the "symmetric" case when $b = a = 1$, the bound in (1.1) (with $\tau = \tau_4$) reduces to $1.34/\sqrt{n}$, which is $\frac{25}{1.34} > 18$ times as small as the bound in (1.10) $\big($for $n \geqslant (2b)^2 = 4\big)$.

While the two-point distributions may be of particular interest, they are of a bounded support set, and hence all their moments are finite. On the other hand, one may object that the bounds given in Theorems 1.1 and 1.2 will be infinite and hence useless if the 4th-order moments of the X_i's are infinite. However, this concern is easily addressed via truncation.

For a minute, let X denote any zero-mean r.v. If the distribution of X is continuous, then for each $b \in [0, \infty]$ there is some $a \in [0, \infty]$ such that the r.v. $X^{a,b} := X\,\mathbf{I}\{-a < X < b\}$ is zero-mean; the same holds in the case when the distribution of X is symmetric (about 0) — then one can simply take $a = b$. If the zero-mean distribution of X is not continuous or symmetric, one can use randomization, say as in [15], to still find, for each $b \in [0, \infty]$, some $a \in [0, \infty]$ and some zero-mean r.v. $X^{a,b}$ such that $\mathsf{P}(-a \leqslant X^{a,b} \leqslant b) = 1$ and $X^{a,b} = X$ on the event $\{-a < X < b\}$; let us refer to any such r.v. $X^{a,b}$ as a zero-mean truncation of the zero-mean r.v. X. $\big($One could similarly base an appropriate construction on the so-called Winsorization $(-a) \vee (X \wedge b)$ instead of the truncation $X\,\mathbf{I}\{-a < X < b\}$.$\big)$

Now let $X_1, \dots, X_n$ be zero-mean r.v.'s as in Theorem 1.1 or 1.2. Respectively, let $B(X_1, \dots, X_n)$ denote (for any of the triples $\tau_1, \dots, \tau_4, \tilde{\tau}_{1,1}, \dots, \tilde{\tau}_{4,1}$), either one of the bounds in (1.1) or (1.3), as it depends on (the individual distributions of) the X_i's. So, $B(X_1, \dots, X_n)$ denotes the bound in (1.1) under the conditions of Theorem 1.1, and it denotes the bound in (1.3) under the conditions of Theorem 1.2. The following corollary of Theorems 1.1 and 1.2 is immediate:

Corollary 1.5. *Under the conditions of Theorem 1.1 or 1.2, for each $i \in \{1, \dots, n\}$ let $X_i^{a_i, b_i}$ be a zero-mean truncation of X_i. Then for all $z \in \mathbb{R}$*

$$|\,\mathsf{P}(T \leqslant z) - \Phi(z)| \leqslant \mathsf{P}\Big(\bigcup_1^n \{X_i \notin (-a_i, b_i)\} \Big) + B(X_1^{a_1, b_1}, \dots, X_n^{a_n, b_n}). \quad (1.11)$$

Note that the upper bound in (1.11) can be expressed only in terms of the individual distributions of the X_i's (rather than their joint distribution), since

$$\mathsf{P}\Big(\bigcup_1^n \{X_i \notin (-a_i, b_i)\} \Big) = 1 - \prod_1^n \mathsf{P}\big(X_i \in (-a_i, b_i) \big).$$

So, when the bound in (1.1), (1.3), (1.6), or (1.10) can be computed, usually the "truncated" bound in (1.11) can be computed as well.

One may want to compare the bound in (1.11) with that in (1.10) or even with the "truncated" version of the latter bound:

$$1 - \prod_1^n \mathsf{P}\left(X_i \in (-a_i, b_i)\right) + 10.2\tilde{\gamma}_2 + 25\tilde{\gamma}_3, \tag{1.12}$$

where $\tilde{\gamma}_2$ and $\tilde{\gamma}_3$ are obtained from γ_2 and γ_3 by replacing the X_i's with their zero-mean truncations $X_i^{a_i, b_i}$, as in Corollary 1.5.

Let us make such a comparison when the X_i's are i.i.d. with a common distribution, which is either the Student distribution with $d > 0$ degrees of freedom or the (centered) Pareto distribution with the density

$$f_s(x) := s\left(x + \frac{s}{s-1}\right)^{-s-1} \mathbf{I}\left\{x > -\frac{1}{s-1}\right\},$$

where s is a parameter with values in the interval $(1, \infty)$. Clearly, Student's distribution with d degrees of freedom is symmetric, with heavy tails for small d and light ones for large d, whereas the Pareto distribution with parameter s is highly skewed to the right, with a heavy right tail for small $s > 1$ and a light one for large s. In keeping with the "i.i.d." assumption, let us consider the "truncated" bounds in (1.11) and (1.12) with $b_1 = \cdots = b_n =: b$ and, accordingly, $a_1 = \cdots = a_n =: a$; note that in each of the two cases under consideration (Student's or Pareto's), the value of a is uniquely determined by that of b. Then, moreover, let us (numerically) minimize the "truncated" bounds in b. The results are shown in Figures 2 and 3. There, the graphs are shown: of the bound in (1.10) (blue), of the minimized "truncated" bound (1.12) (magenta), of the bound in (1.1) (red), and of the minimized "truncated" bound in (1.11) (green) — for sample sizes $n \in \{10, 100, 1000, 10000\}$, $d \in [2.5, 20]$, and $s \in [3.5, 20]$; at that, for the "red" and "green" bounds the triple $\tau_2 = (2.01, 1.02, 0.61)$ of constant factors in (1.1) is used.
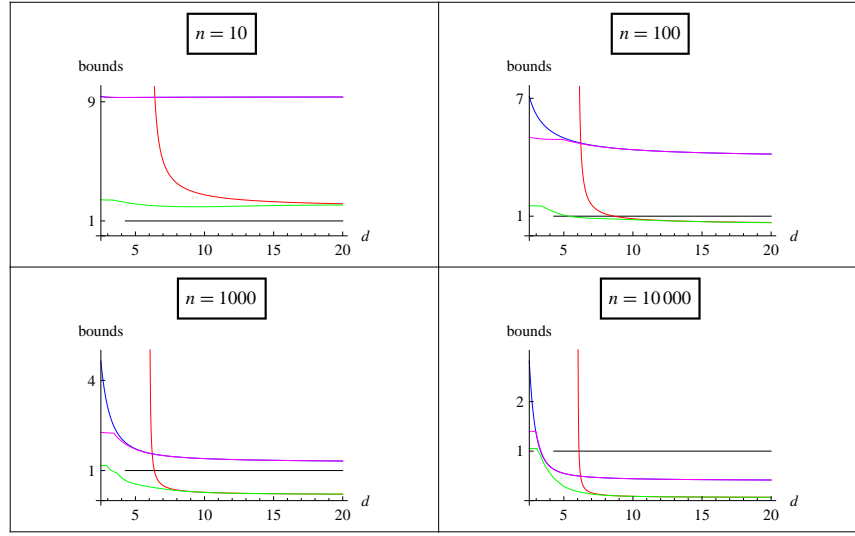
FIG 2. *The bounds in the case of Student's distribution with d degrees of freedom.*
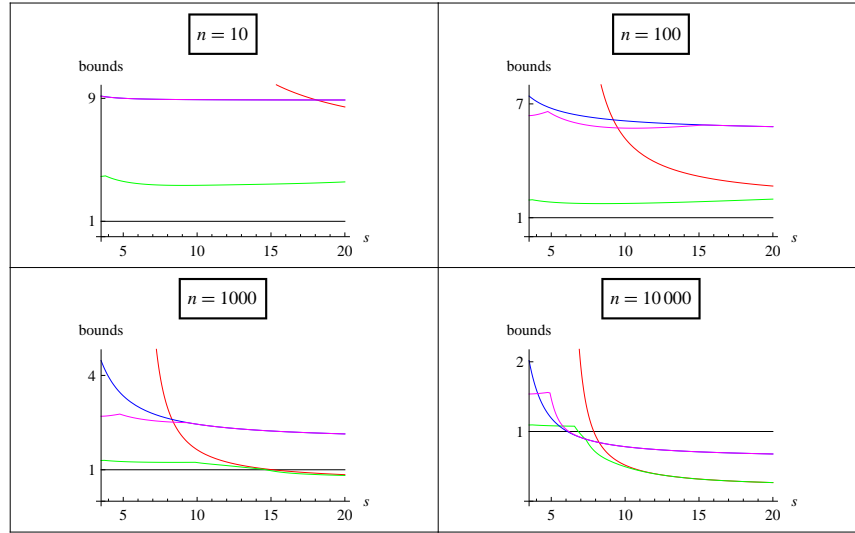
FIG 3. *The bounds in the case of Pareto's distribution with parameter s.*

These pictures suggest the following.

1. Predictably, truncation helps significantly only when the tails are heavy enough — that is, for small enough values of the parameters d and s. Predictably as well, truncation is much more useful with the bound in

(1.1) than it is with that in (1.10).

2. For Student's and Pareto's distributions, even the minimized "truncated" bound in (1.12) is nontrivial (that is, less than 1) only if n is greater than 1000 (or even a few thousands). In fact, this bound is not much less than 0.5 even for $n = 10000$ and light tails. For instance, for $n = 10000$ and Student's distribution with $d = 20$ d.f., the bound in (1.10) and the minimized bound in (1.12) are both ≈ 0.417, whereas the bound in (1.1) and the minimized bound in (1.11) are both ≈ 0.068 (again, with $\tau = \tau_2 = (2.01, 1.02, 0.61)$).

3. Figure 3, for the Pareto case, as as well as other considerations (see e.g. [14, 15] and discussion therein) suggest that the Student statistic may not be appropriate for statistical inference when the underlying distribution is significantly skewed. Alternative statistics, "correcting" for the asymmetry, were offered and considered; see [14, 15] and discussion therein.

4. If the tails are very heavy, then even the minimized "truncated", "green" bound in (1.11) is not much less than 1 even if n is as large as 1000 and the underlying distribution is symmetric. This may be in broadly considered agreement with the fact, established in [5], that if the the underlying distribution is in the domain of attraction of a stable law with index $\alpha < 2$, then the limit distribution of the self-normalized sum and, equivalently, that of the Student statistic is not normal.

5. For almost all considered values of n, d, and s, the minimized "truncated" bound in (1.11) is significantly less than that in (1.12), except in the Pareto case with $n = 10000$ for a rather short interval of values s near 7, where, however, even the better bound is only slightly less than 1. Conceivably, this deficiency might be fixed by using another triple of constants in place of the triple $\tau_2 = (2.01, 1.02, 0.61)$. Moreover, when the tails are light enough, even the "non-truncated" bound in (1.1) significantly improves both on the "truncated" and "non-truncated" bounds in (1.12) and (1.10). Thus, especially with the truncation tool, getting smaller constant factors may be more effective than insisting on the optimal order of moments even for the price of much greater constants.

It appears that, with the much smaller constant factors than in the preceding results, the bounds presented above may be approaching the state of being of use in statistical practice. There are additional resources to be tapped on. For instance, the proofs of Theorems 1.1 and 1.2 rely to a large extent on a hybrid between the Chebyshev and Cantelli bounds, developed in [9] specifically for the purposes of the present paper. One can similarly try to use and/or develop the much more accurate (but also much more complicated) upper bounds on large deviation probabilities given and discussed in [11]; however, at that the proofs can be expected to be much harder to produce or read.

2. Proofs

Proof of Proposition 1.4. Introduce $\Lambda(a, z) := \Phi(u_{a,z})$, where $a \in (0,1)$ and $u_{a,z} := \frac{z}{\sqrt{1+a(z^2-1)}}$, so that $\Phi_n(z) = \Lambda(\frac{1}{n}, z)$ and $\Phi(z) = \Lambda(0, z)$. By the mean value theorem, for some $b = b_z \in (0, a)$

$$\frac{\Lambda(a, z) - \Lambda(0, z)}{a} = \frac{\partial \Lambda}{\partial a}(b, z) = \frac{u_{b,z}(1 - u_{b,z}^2)\varphi(u_{b,z})}{2(1 - b)},$$

where φ is the standard normal density function. So, to prove inequality (1.8), it suffices to note that $\sup_{u \in \mathbb{R}} |u(1 - u^2)\varphi(u)| = 2C$ and $\frac{1}{1-b} < \frac{1}{1-a} = \frac{n}{n-1}$ for $b \in (0, a)$ and $a = \frac{1}{n}$. That the constant factor C is the best possible in (1.8) follows because, by l'Hospital's rule, $\frac{\Lambda(a,z)-\Lambda(0,z)}{a} \sim \frac{\partial \Lambda}{\partial a}(a, z)$ as $a \downarrow 0$. $\qquad\square$

The proof of Theorem 1.1 is based, in part, on the following two lemmas.

Lemma 2.1. *Take any λ, r_*, a, b in $(0, \infty)$. Take any c and r in $(0, \infty)$ such that*

$$c \geqslant \frac{\lambda}{r} \quad and \quad r \leqslant r_*.$$

Let Y by any r.v. such that $\mathsf{E}\, Y = 0$ and $\sigma := \sqrt{\mathsf{E}\, Y^2} \in (0, \infty)$. Then

$$\mathsf{P}(Y \geqslant c) \leqslant \psi\left(r_*, \frac{\lambda}{\sigma}\right)r, \quad where \quad \psi(u, v) := \frac{u \wedge v}{v^2 + (u \wedge v)^2}. \tag{2.1}$$

Also,

$$\mathsf{P}\left(Y \notin (-a, b)\right) \leqslant \frac{4\sigma^2 + (a - b)^2}{(a + b)^2}. \tag{2.2}$$

Proof of Lemma 2.1. By the condition $c \geqslant \frac{\lambda}{r}$ and Cantelli's inequality,

$$\mathsf{P}(Y \geqslant c) \leqslant \mathsf{P}(Y \geqslant \tfrac{\lambda}{r}) \leqslant \frac{\sigma^2}{\sigma^2 + (\lambda/r)^2} = \frac{r^2}{r^2 + v^2}, \quad where \quad v := \lambda/\sigma.$$

Note that $\frac{r}{r^2+v^2}$ increases in $r \in [0, v]$ and decreases in $r \in [v, \infty)$. So, if $r_* \leqslant v$, then the condition $r \leqslant r_*$ implies $\frac{r}{r^2+v^2} \leqslant \frac{r_*}{r_*^2+v^2} = \psi(r_*, v)$. If now $r_* \geqslant v$, then $\frac{r}{r^2+v^2} \leqslant \frac{v}{v^2+v^2} = \psi(r_*, v)$, so that the inequality in (2.1) holds in this case as well. As for inequality (2.2), it is given in [9]. $\qquad\square$

Lemma 2.2. *For any positive real numbers x, x_1, x_2 such that $x \geqslant x_1 \vee x_2$, one has*

$$x_1\overline{\Phi}(x) \leqslant \overline{\Phi}^*(x_2), \tag{2.3}$$

where

$$\overline{\Phi} := 1 - \Phi,$$
$$\overline{\Phi}^*(x) := 0.17\,\mathbf{I}\{0 < x < 0.752\} + x\overline{\Phi}(x)\,\mathbf{I}\{x \geqslant 0.752\}.$$

Proof of Lemma 2.2. It is well-known that the function $\overline{\Phi}$ is log-concave; see e.g. [4, 12]. So, the function L defined on $(0,\infty)$ by the formula $L(x) := \ln\big(x\overline{\Phi}(x)\big)$ is concave, and hence $L(x) \leqslant L(x_0) + L'(x_0)(x - x_0)$ for any x and x_0 in $(0,\infty)$. Also, $L'(0.751) > 0 > L'(0.752)$ and $L(0.752) + L'(0.752)(0.751 - 0.752) < \ln 0.17$. This implies that $L < \ln 0.17$ on $(0,\infty)$ and L is decreasing on $(0.752, \infty)$, whence $\sup\{x\overline{\Phi}(x)\colon x \geqslant z\} \leqslant \overline{\Phi}^*(z)$ for all $z \in (0,\infty)$. Now the lemma follows. $\square$

Proof of Theorem 1.1. This proof uses some of the ideas in the proof of (1.9) in [6], which were previously presented in [7, 8]. As mentioned before, there are a number of mistakes of various kinds in the proof in [6]. For instance (in the notations of [6]), a bound on $|\Phi(\frac{(1-\varepsilon)r\sigma}{\sqrt{n}\,\sigma_n(r)}) - \Phi(\frac{(1-\varepsilon)r}{\sqrt{n}})|$ analogous to that on $|\Phi(\frac{r\sigma}{\sqrt{n}\,\sigma_n(r)}) - \Phi(\frac{r}{\sqrt{n}})|$ in [6, (1.12)] is missing there; moreover, the same bound in [6, (1.12)] must have $(\frac{\sigma}{\sigma_n(r)} \wedge 1)^2$ instead of $(\frac{\sigma}{\sigma_n(r)} \wedge 1)$. We have also produced and utilized some new ideas in this proof. One of them is presented in Lemma 2.1 above, which depends on the result of [9], specifically developed for the purposes of the present paper.

Without loss of generality, assume that

$$\beta_2 = 1.$$

Take any

$$\kappa \in (0,\infty), \quad \varepsilon_4 \in (0, \tfrac{1}{2}), \quad \varepsilon_3 \in (0,\infty), \quad \varepsilon_2 \in (0,1),$$
$$\theta_3 \in (0,1), \quad \theta_4 \in (0,\infty) \tag{2.4}$$

and introduce

$$\Delta := \Delta(z) := \mathsf{P}(T \leqslant z) - \Phi(z) \tag{2.5}$$

and also

$$r_3 := \beta_3, \quad r_4 := \tilde{\beta}_4^{1/2}, \quad r_6 := \frac{\tilde{\beta}_6}{\beta_3^3}, \tag{2.6}$$

$$\varepsilon := \kappa r_4, \quad \tilde{\varepsilon}_4 := \frac{\varepsilon_4}{\kappa}. \tag{2.7}$$

It suffices to show that

$$|\Delta| = |\Delta(z)| \leqslant A_3 r_3 + A_4 r_4 + A_6 r_6,$$

where without loss of generality let us assume that

$$z > 0.$$

Consider the following three cases.
Case 1 ("small n"): $\varepsilon \geqslant \varepsilon_4$ *or* $r_3 \geqslant \varepsilon_3$. Note that

$$\varepsilon \geqslant \varepsilon_4 \iff r_4 \geqslant \tilde{\varepsilon}_4.$$

So,

$$|\Delta| \leqslant 1 \leqslant (A_{3,1}\, r_3) \vee (A_{4,1}\, r_4) \vee (A_{6,1}\, r_6)$$
$$\leqslant A_{3,1}\, r_3 + A_{4,1}\, r_4 + A_{6,1}\, r_6, \quad \text{where} \qquad (2.8)$$
$$A_{3,1} := \frac{1}{\varepsilon_3}, \quad A_{4,1} := \frac{1}{\tilde{\varepsilon}_4}, \quad A_{6,1} := 0.$$

Case 2 ("large n" & "large deviations"): $\varepsilon < \varepsilon_4$ & $r_3 < \varepsilon_3$ & $z \geqslant \frac{\theta_3}{r_3} \wedge \frac{\theta_4}{r_4}$. Then, by (2.5) and (2.1),

$$|\Delta| \leqslant (P_1 + P_2) \vee \overline{\Phi}(z),$$

where

$$P_1 := \mathsf{P}\left(T > z, V > 1 - \varepsilon_2\right)$$
$$\leqslant \mathsf{P}(S > (1 - \varepsilon_2)z) \leqslant \left(\psi(\varepsilon_3, \tilde{\theta}_3)r_3\right) \vee \left(\psi(\tilde{\varepsilon}_4, \tilde{\theta}_4)r_4\right),$$
$$\tilde{\theta}_j := (1 - \varepsilon_2)\theta_j,$$
$$P_2 := \mathsf{P}\left(V \leqslant 1 - \varepsilon_2\right) = \mathsf{P}\left(\sum_1^n (\mathsf{E}\, X_i^2 - X_i^2) \geqslant \tilde{\varepsilon}_2\right) \leqslant \psi(\tilde{\varepsilon}_4, \tilde{\varepsilon}_2)r_4,$$
$$\tilde{\varepsilon}_2 := \varepsilon_2(2 - \varepsilon_2).$$

Note also that the currently assumed case conditions $\varepsilon < \varepsilon_4$ & $r_3 < \varepsilon_3$ & $z > \frac{\theta_3}{r_3} \wedge \frac{\theta_4}{r_4}$ imply $z > \frac{\theta_3}{r_3} > \frac{\theta_3}{\varepsilon_3}$ or $z > \frac{\theta_4}{r_4} > \frac{\theta_4}{\tilde{\varepsilon}_4}$. So, Lemma 2.2 yields

$$\overline{\Phi}(z) \leqslant \left[\overline{\Phi}^*\left(\tfrac{\theta_3}{\varepsilon_3}\right) \tfrac{r_3}{\theta_3}\right] \vee \left[\overline{\Phi}^*\left(\tfrac{\theta_4}{\tilde{\varepsilon}_4}\right) \tfrac{r_4}{\theta_4}\right].$$

Thus,

$$|\Delta| \leqslant A_{3,2}\, r_3 + A_{4,2}\, r_4 + A_{6,2}\, r_6, \quad \text{where} \qquad (2.9)$$
$$A_{3,2} := \psi(\varepsilon_3, \tilde{\theta}_3) \vee \left[\overline{\Phi}^*\left(\tfrac{\theta_3}{\varepsilon_3}\right) \tfrac{1}{\theta_3}\right],$$
$$A_{4,2} := [\psi(\tilde{\varepsilon}_4, \tilde{\theta}_4) + \psi(\tilde{\varepsilon}_4, \tilde{\varepsilon}_2)] \vee \left[\overline{\Phi}^*\left(\tfrac{\theta_4}{\tilde{\varepsilon}_4}\right) \tfrac{1}{\theta_4}\right],$$
$$A_{6,2} := 0.$$

Case 3 ("large n" & "moderate deviations"): $\varepsilon < \varepsilon_4$ & $r_3 < \varepsilon_3$ & $z < \frac{\theta_3}{r_3} \wedge \frac{\theta_4}{r_4}$. In this case, note that

$$\{T \leqslant z\} = \{T_z \leqslant z\}, \quad \text{where} \quad T_z := S - z\left(\sqrt{1 + \eta} - 1\right), \quad \text{and} \quad \eta := V^2 - 1.$$

Note also that the expression $S - z\left(\sqrt{1 + \eta} - 1\right)$ for T_z is convex in (S, η), so that its linear approximation $\left(\text{at the point } (\mathsf{E}\,S, \mathsf{E}\,\eta) = (0, 0)\right)$

$$S_z := S - z\eta/2$$

never exceeds T_z, whence

$$\delta := \frac{T_z - S_z}{z} = 1 + \frac{\eta}{2} - \sqrt{1 + \eta} \geqslant 0.$$

Therefore and because $\mathsf{P}(T \leqslant z) = \mathsf{P}(T_z \leqslant z)$, one has

$$\mathsf{P}\left(S_z \leqslant (1-\varepsilon)z\right) - \mathsf{P}(\delta > \varepsilon) \leqslant \mathsf{P}(T \leqslant z) \leqslant \mathsf{P}(S_z \leqslant z).$$

In view of (2.5), it follows that

$$\Delta \leqslant \mathrm{BE} + D(1) \quad \text{and}$$
$$-\Delta \leqslant \mathrm{BE} + \mathsf{P}(\delta > \varepsilon) + D(1-\varepsilon) + \tilde{D}_\varepsilon, \quad \text{where}$$

$$\mathrm{BE} := \sup_{u \in \mathbb{R}} \left| \mathsf{P}\left(S_z \leqslant u\right) - \Phi\left(\frac{u}{\sigma_z}\right) \right|,$$

$$\sigma_z := \sqrt{\mathsf{E}\,S_z^2} = \sqrt{\sum_1^n \mathsf{E}\,X_{i,z}^2},$$

$$D(u) := \left| \Phi(uz) - \Phi\left(\frac{uz}{\sigma_z}\right) \right|, \tag{2.10}$$

$$\tilde{D}_\varepsilon := \sup_{x \geqslant 0} \left[\Phi(x) - \Phi\left((1-\varepsilon)x\right) \right]. \tag{2.11}$$

Thus,

$$|\Delta| \leqslant \mathrm{BE} + \mathsf{P}(\delta > \varepsilon) + D(1) \vee D(1-\varepsilon) + \tilde{D}_\varepsilon. \tag{2.12}$$

Note also that

$$S_z = \sum_1^n X_{i,z}, \quad \text{where}$$

$$X_{i,z} := X_i - zY_i/2 \quad \text{and} \quad Y_i := X_i^2 - \mathsf{E}\,X_i^2, \quad \text{whence}$$

$$\eta = \sum_1^n Y_i.$$

By a recent result of Shevtsova [17],

$$\mathrm{BE} \leqslant 0.56\,\frac{\beta_{3,z}}{\sigma_z^3}, \tag{2.13}$$

where

$$\beta_{3,z} := \sum_1^n \mathsf{E}\,|X_{i,z}|^3 \leqslant \sum_1^n \mathsf{E}\left(|X_i| + \frac{z}{2}|Y_i|\right)^3 \leqslant \frac{\beta_3}{(1-\alpha)^2} + \left(\frac{z}{2}\right)^3 \frac{\tilde{\beta}_6}{\alpha^2}, \tag{2.14}$$

for any

$$\alpha \in (0,1); \tag{2.15}$$

the second inequality in (2.14) follows from the elementary inequality $(a+b)^3 \leqslant \frac{a^3}{(1-\alpha)^2} + \frac{b^3}{\alpha^2}$ for all a and b in $[0,\infty)$ and $\alpha \in (0,1)$. Recalling also the condition $z < \frac{\theta_3}{r_3} \wedge \frac{\theta_4}{r_4}$ and definitions (2.6), one has

$$\beta_{3,z} \leqslant \frac{1}{(1-\alpha)^2}\,r_3 + \frac{\theta_3^3}{8\alpha^2}\,r_6. \tag{2.16}$$

Next,

$$\sigma_z^2 = \sum_1^n \mathsf{E}\, X_{i,z}^2 = 1 + \left(\frac{z}{2}\right)^2 \tilde{\beta}_4 - z \sum_1^n \mathsf{E}\, X_i^3 \geqslant 1 - z\beta_3 > 1 - \theta_3. \qquad (2.17)$$

So, (2.13) and (2.16) yield

$$\mathrm{BE} \leqslant \frac{0.56}{(1-\theta_3)^{3/2}}\left(\frac{1}{(1-\alpha)^2}\, r_3 + \frac{\theta_3^3}{8\alpha^2}\, r_6\right). \qquad (2.18)$$

Further, since $0 < \varepsilon < \varepsilon_4 < \frac{1}{2}$, one has $\delta > \varepsilon \iff \eta \notin [2\varepsilon - 2\sqrt{2\varepsilon}, 2\varepsilon + 2\sqrt{2\varepsilon}]$. So, by (2.2), (2.6), and (2.7),

$$\mathsf{P}(\delta > \varepsilon) \leqslant \frac{4r_4^2 + 16\varepsilon^2}{32\varepsilon} = \frac{1 + 4\kappa^2}{8\kappa}\, r_4. \qquad (2.19)$$

Next, by (2.10), for any $u \in [0, \infty)$,

$$D(u) \leqslant uz\left|\frac{1}{\sigma_z} - 1\right|\varphi\left(\frac{uz}{\sigma_z \vee 1}\right), \qquad (2.20)$$

where φ is the standard normal density function. By the equalities in (2.17) and the case conditions $\varepsilon < \varepsilon_4$ and $z < \frac{\theta_3}{r_3} \wedge \frac{\theta_4}{r_4}$,

$$|\sigma_z^2 - 1| \leqslant z\beta_3 + \left(\frac{z}{2}\right)^2 \tilde{\beta}_4 = zr_3 + \left(\frac{z}{2}\right)^2 r_4^2 \leqslant zr_3 + \left(\frac{z}{2}\right)^2 \tilde{\varepsilon}_4 r_4 \qquad (2.21)$$

and

$$|\sigma_z^2 - 1| \leqslant zr_3 + \left(\frac{z}{2}\right)^2 r_4^2 \leqslant \theta_3 + \theta_4^2/4. \qquad (2.22)$$

Writing $|\frac{1}{\sigma_z} - 1| = \frac{|\sigma_z^2 - 1|}{\sigma_z + \sigma_z^2}$, and using (2.20) and (2.21), one has

$$D(u) \leqslant D_1(u) + D_2(u), \quad \text{where}$$

$$D_1(u) := r_3\, \frac{v^2\varphi(v)}{u}\, \rho_2, \quad D_2(u) := r_4\, \frac{\tilde{\varepsilon}_4}{4}\, \frac{v^3\varphi(v)}{u^2}\, \rho_3,$$

$$v := \frac{uz}{\sigma_z \vee 1}, \quad \rho_j := \frac{(\sigma_z \vee 1)^j}{\sigma_z + \sigma_z^2}.$$

If $\sigma_z \leqslant 1$, then by (2.17) for $j = 2, 3$

$$\rho_j = \frac{1}{\sigma_z + \sigma_z^2} \leqslant \rho_* := \frac{1}{1 - \theta_3 + \sqrt{1 - \theta_3}}.$$

If $\sigma_z > 1$, then by (2.22) for $j = 2, 3$

$$\rho_j = \frac{\sigma_z^j}{\sigma_z + \sigma_z^2} = \frac{1}{\sigma_z^{1-j} + \sigma_z^{2-j}} \leqslant \rho_{**,j} := \frac{1}{\sigma_*^{1-j} + \sigma_*^{2-j}},$$

where

$$\sigma_* := \sqrt{1 + \theta_3 + \theta_4^2/4}.$$

Note also that

$$\sup_{v>0} v^j \varphi(v) = s_j := \frac{1}{\sqrt{2\pi}} \left(\frac{j}{e}\right)^{j/2}$$

for $j = 2, 3$. Therefore, recalling also the condition $\varepsilon < \varepsilon_4$, one has

$$D(1) \vee D(1 - \varepsilon) \leqslant r_3 \frac{s_2}{1 - \varepsilon_4} (\rho_* \vee \rho_{**,2}) + r_4 \frac{\tilde{\varepsilon}_4}{4} \frac{s_3}{(1 - \varepsilon_4)^2} (\rho_* \vee \rho_{**,3}). \quad (2.23)$$

Next, let us estimate $\tilde{D}_\varepsilon$. First here, one can use a special-case l'Hospital-type rule for monotonicity, such as [13, Proposition 4.1], to see that for each $x \in (0, \infty)$ the ratio $\frac{\Phi(x) - \Phi((1-t)x)}{t}$ increases in $t \in (0, 1)$. On the other hand, for each $t \in (0, 1)$ the expression $\Phi(x) - \Phi\big((1 - t)x\big)$ attains its maximum in $x \in (0, \infty)$ at $x = x_t$, where

$$x_t := \sqrt{-\frac{2\ln(1 - t)}{t(2 - t)}}.$$

On recalling also the definition (2.11) of $\tilde{D}_\varepsilon$ and the conditions $0 < \varepsilon < \varepsilon_4 < \frac{1}{2}$, it follows that

$$\tilde{D}_\varepsilon \leqslant R(\varepsilon_4)\varepsilon = R(\varepsilon_4)\kappa r_4, \quad \text{where} \quad R(\varepsilon_4) := \frac{\Phi(x_{\varepsilon_4}) - \Phi\big((1 - \varepsilon_4)x_{\varepsilon_4}\big)}{\varepsilon_4}. \quad (2.24)$$

Collecting (2.12), (2.18), (2.19), (2.23), and (2.24), one bounds $|\Delta|$ in Case 3 as follows:

$$|\Delta| \leqslant A_{3,3}\, r_3 + A_{4,3}\, r_4 + A_{6,3}\, r_6, \quad \text{where} \quad (2.25)$$

$$A_{3,3} := \frac{0.56}{(1 - \theta_3)^{3/2}(1 - \alpha)^2} + \frac{s_2(\rho_* \vee \rho_{**,2})}{1 - \varepsilon_4},$$

$$A_{4,3} := \frac{1 + 4\kappa^2}{8\kappa} + \frac{\tilde{\varepsilon}_4 s_3 (\rho_* \vee \rho_{**,3})}{4(1 - \varepsilon_4)^2} + R(\varepsilon_4)\kappa,$$

$$A_{6,3} := \frac{0.07}{\alpha^2} \left(\frac{\theta_3^2}{1 - \theta_3}\right)^{3/2}.$$

Collecting now the bounds (2.8), (2.9), and (2.25) on $|\Delta|$ in Cases 1–3, one concludes that in all of the three cases

$$|\Delta| \leqslant A_3\, r_3 + A_4\, r_4 + A_6\, r_6, \quad \text{where} \quad (2.26)$$

$$A_p := A_{p,1} \vee A_{p,2} \vee A_{p,3}$$

for $p = 3, 4, 6$.

Now one can arbitrarily select positive "weights" w_3, w_4, w_6 and then try numerical minimization of (say) $(w_3 A_3) \vee (w_4 A_4) \vee (w_6 A_6)$ with respect to all

the parameters: α, ε_4, ε_3, ε_2, κ, θ_3, θ_4, within their specified ranges — recall (2.4) and (2.15). The target function here appears to have a great number of local minima, and so, it is hardly possible to find the global minimum. Even though the numerical minimization is imperfect, it should be clear that the bound in (2.26) holds for all the allowable values of the parameters as specified in (2.4). The following table shows the values of the parameters α, ε_4, ε_3, ε_2, κ, θ_3, θ_4 found by the mentioned numerical minimization for each of a few selected triples (w_3, w_4, w_6), as well as the resulting triple τ_i of the coefficients (A_3, A_4, A_6), corresponding to the so obtained values of the parameters.

w_3	w_4	w_6	α	ε_4	ε_3	ε_2	κ	θ_3	θ_4	triple
1	1	1	$\frac{2}{25}$	$\frac{123}{10^3}$	$\frac{2703}{10^3}$	$\frac{22}{125}$	$\frac{43}{250}$	$\frac{377}{10^3}$	$\frac{5407}{10^3}$	τ_1
1	2	1	$\frac{27}{200}$	$\frac{363}{10^3}$	$\frac{1401}{10^3}$	$\frac{19}{50}$	$\frac{91}{250}$	$\frac{413}{10^3}$	$\frac{3167}{10^3}$	τ_2
1	1	10^6	$\frac{381}{500}$	$\frac{471}{10^3}$	$\frac{6927}{10^3}$	$\frac{23}{10^3}$	$\frac{79}{50}$	$\frac{9}{200}$	$\frac{3809}{10^3}$	τ_3
1	10^{-5}	10^{-6}	$\frac{8.39}{10^5}$	$\frac{3.17}{10^5}$	1.32	$\frac{3.49}{10^5}$	$\frac{9.97}{10^7}$	0.3738	2.69	τ_4

Now Theorem 1.1 is completely proved. □

Proof of Theorem 1.2. This proof is quite similar to that of Theorem 1.1. The only essential difference that, instead of the constant 0.56 in (2.13) one can now use the better constant 0.4785, according to a recent result of Tyurin [18]. Because we cannot find the global minima, it sometimes turns out that the numerical minimization with the better constant 0.4785 produces results worse (or not quite better) than those obtained using the worse constant 0.56. (!) In such cases, we used the values of the parameters α, ε_4, ε_3, ε_2, κ, θ_3, θ_4 found in the general, non-iid setting — with the worse constant 0.56 and with the same weights (w_3, w_4, w_6); the resulting triples are denoted as $\tilde{\tau}_{i,2}$, with the second subscript 2. Otherwise, the triple's second subscript is 1, as in $\tilde{\tau}_{1,1}$, $\tilde{\tau}_{3,1}$, and $\tilde{\tau}_{4,1}$. See the table below.

w_3	w_4	w_6	α	ε_4	ε_3	ε_2	κ	θ_3	θ_4	triple
1	1	1	$\frac{41}{500}$	$\frac{113}{500}$	$\frac{277}{100}$	$\frac{39}{200}$	$\frac{83}{500}$	$\frac{409}{10^3}$	$\frac{4467}{10^3}$	$\tilde{\tau}_{1,1}$
1	1	1	$\frac{2}{25}$	$\frac{123}{10^3}$	$\frac{2703}{10^3}$	$\frac{22}{125}$	$\frac{43}{250}$	$\frac{377}{10^3}$	$\frac{5407}{10^3}$	$\tilde{\tau}_{1,2}$
1	2	1	$\frac{27}{200}$	$\frac{363}{10^3}$	$\frac{1401}{10^3}$	$\frac{19}{50}$	$\frac{91}{250}$	$\frac{413}{10^3}$	$\frac{3167}{10^3}$	$\tilde{\tau}_{2,2}$
1	2.1	1	0.14	$\frac{275}{10^3}$	6.7	0.42	0.27	0.44	3.2	$\tilde{\tau}_{2.1,1}$
1	1	10^6	$\frac{777}{10^3}$	$\frac{1}{2}$	$\frac{1381}{500}$	$\frac{27}{10^3}$	$\frac{451}{10^3}$	$\frac{47}{10^3}$	$\frac{4569}{500}$	$\tilde{\tau}_{3,1}$
1	10^{-5}	10^{-6}	$\frac{3}{10^4}$	$\frac{43}{10^5}$	10.3	$\frac{13}{10^4}$	3.5	0.401	1.6	$\tau_{4,1}$

For instance, one can see that the values of the parameters α, ε_4, ε_3, ε_2, κ, θ_3, θ_4 resulting in the triple $\tilde{\tau}_{1,2}$ are the same those used to obtain the triple τ_1. Similarly, the values of the parameters for the triple $\tilde{\tau}_{2,2}$ are the same those for the triple τ_2. □

References

[1] V. Bentkus, M. Bloznelis, and F. Götze. A Berry-Esséen bound for Student's statistic in the non-i.i.d. case. *J. Theoret. Probab.*, 9(3):765–796, 1996.

[2] V. Bentkus and F. Götze. The Berry-Esseen bound for Student's statistic. *Ann. Probab.*, 24(1):491–503, 1996.

[3] S. Bourguin and C. Tudor. Malliavin calculus and self normalized sums, preprint, arxiv:1107.0410v1 [math.PR].

[4] R. L. Hall, M. Kanter, and M. D. Perlman. Inequalities for the probability content of a rotated square and related convolutions. *Ann. Probab.*, 8(4):802–813, 1980.

[5] B. F. Logan, C. L. Mallows, S. O. Rice, and L. A. Shepp. Limit distributions of self-normalized sums. *Ann. Probability*, 1:788–809, 1973.

[6] S. V. Nagaev. The Berry-Esseen bound for self-normalized sums. *Siberian Adv. Math.*, 12(3):79–125 (2003), 2002.

[7] S. Y. Novak. On self-normalized sums. *Math. Methods Statist.*, 9(4):415–436, 2000.

[8] S. Y. Novak. On self-normalized sums of random variables and the Student's statistic. *Theory Probab. Appl.*, 49(2):336–344, 2005.

[9] I. Pinelis. Between Chebyshev and Cantelli (preprint), arXiv:1011.6065 [math.PR].

[10] I. Pinelis. Exact bounds on the closeness between the Student and standard normal distributions, preprint, http://arxiv.org/find/all/1/au:+pinelis/0/1/0/all/0/1.

[11] I. Pinelis. On the Bennett-Hoeffding inequality (preprint), arXiv:0902.4058v1 [math.PR].

[12] I. Pinelis. Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities. In *Advances in stochastic inequalities (Atlanta, GA, 1997)*, volume 234 of *Contemp. Math.*, pages 149–168. Amer. Math. Soc., Providence, RI, 1999.

[13] I. Pinelis. On l'Hospital-type rules for monotonicity. *JIPAM. J. Inequal. Pure Appl. Math.*, 7(2):Article 40, 19 pp. (electronic), 2006.

[14] I. Pinelis. Exact inequalities for sums of asymmetric random variables, with applications. *Probab. Theory Related Fields*, 139(3-4):605–635, 2007.

[15] I. Pinelis. Optimal two-value zero-mean disintegration of zero-mean random variables. *Electron. J. Probab.*, 14:no. 26, 663–727, 2009.

[16] Q.-M. Shao. An explicit Berry-Esseen bound for Student's t-statistic via Stein's method. In *Stein's method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 143–155. Singapore Univ. Press, Singapore, 2005.

[17] I. Shevtsova. Refinement of estimates of the rate of convergence in the Lyapunov theorem (preprint, in Russian).

[18] I. Tyurin. New estimates of the convergence rate in the Lyapunov theorem (preprint, arXiv:0912.0726v1 [math.PR]).